Electron group localization in atoms and molecules

Bernard Silvi, Esmaïl Alikhani

Electron group localization in atoms and molecules

Bernard Silvi[1] and Esmaïl Alikhani[2]
[1)] *Sorbonne Universités, UPMC, Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005 Paris, France*[a)]
[2)] *Sorbonne Universités, UPMC, Univ Paris 06, MONARIS, UMR 8233, 4 place Jussieu, F-75005 Paris Cedex 05, France*

(*Electronic mail: sbernard@lct.jussieu.fr)

(Dated: 16 June 2022)

Partitioning atomic and molecular charge densities in non overlapping chemically significant regions is a challenging problem for quantum chemists. The present method aims to build a tool enabling the determination of "good boundaries" with the help of elementary statistical methods or information theory. This is done by minimizing an objective function with respect to the boundaries of the localization regions, the choice of this function being guided by a clarity requirement. With the sum of the indexes of dispersion (ΣD) or the mutual information (MI) as objective function, the method yield partitions in good agreement with the Aufbau rules for Li-Rn atoms and with Lewis's pairing model for molecules.

I. INTRODUCTION

The concept of chemical bond belongs to a chemical representation of the matter at a microscopic level in terms of atoms linked by bonds developed by chemists at the beginning of the XX$^{\text{th}}$ before the advent of Quantum Mechanics. A closer description of the physical reality would consider unambiguously defined interacting quantum particles rather than atoms and bonds. However, as written by S. Alvarez, R. Hoffmann and C. Mealli: "Chemistry has done more than well in creating a universe of structure and function on the molecular level with just this "imperfectly defined" concept of a chemical bond. Or maybe it has done so well precisely because the concept is flexible and fuzzy."[1]. The same authors noticed that the nature of chemical bond continues to be debated. In fact, the epistemological weakness of the chemical bond concept can be viewed as a consequence of a chemical representation aiming to support chemical explanations rather than as a faithful picture of the reality complying demarcation criteria such as falsifiability.

One of the aims of Quantum Chemistry is to build bridges between the realm of Quantum Mechanics and the world of chemical concepts, in other words to find correspondence between mathematically defined objects and chemical concepts. Two principal routes are used to achieved this task. On the one hand are orbital based methods in which the mathematical intermediates enabling the calculation of approximate wave functions acquire a chemical meaning and on the other hand are methods which use a complementary theory to build strictly defined objects. Among the latter methods is the Quantum Theory of Atoms in Molecules (QTAIM) developed by Richard Bader and co-workers[2] in which the dynamical system theory is used to partition the space occupied by a solid or a molecule in basin of attractors of the one electron density gradient field. These basins satisfy the local virial theorem and are called atomic basins. The atom in molecule is then defined as the union of a nucleus and of the one electron density in its atomic basin. QTAIM does not aim to recover the chemical picture and, during his talks, Richard Bader often claimed " there is no bond, there is bonding". It is possible, however, to get closer the chemical representation with other gradient vector fields, such as that of the electron localization function (ELF)[3], used in the same topological framework[4].

The present article is focussed on the spacial distribution of the electron density. In this respect, we propose to achieve a spacial partition of the molecular position space in terms of regions dominated by definite electron groups with the help of techniques requiring *as few as possible assumptions*. These space-filling additive volumes, the electron group localization regions, are required to minimize the fluctuations of the number of electrons within them. Unlike most partition techniques used in Quantum Chemistry which consider the extrema of radial functions to evidence atomic shell structures[5–12], the boundaries of the whole set of electron groups are optimized simultaneously by minimizing an objective function accounting for the fluctuation of electron numbers. This new approach intend to complement, and even to be an alternative to, the ELF population analysis based on the topology of a local function. For all atoms in the Li-Rn range and for a series of small representative molecules there is a satisfactory correspondence between the electron group localization regions and classical chemical objects such as atomic shells, Lewis's bonding and non-bonding electron pairs and Valence Shell Electron Pair Repulsion (VSEPR) electronic domains.

II. OVERVIEW OF THE METHOD

Most of chemical concepts in use in interpretative methods derive from the Lewis bonding model[13,14]. They rely on a description in which electrons are assigned to atomic kernels and atomic valence shells themselves composed of bonding and non bonding pairs. The kernel and the valence shell are brought together to form an atom in the molecule. As the bonding pairs belong to (at least) two valence shells, Lewis's atoms fundamentally differ from QTAIM atoms: the

[a)] https://www.lct.jussieu.fr/pagesperso/silvi/

sum of the atomic population is the number of electrons in the QTAIM approach whereas it exceeds that number by two times the number of bonding pairs in the Lewis model. The VSEPR model considers the arrangement of the bonding and non bonding pairs in the valence shells to explain the molecular geometry[15,16]. Although there is neither experimental or theoretical proofs of the reality of Lewis's pairs and VSEPR electronic domains, these models account for the stoichiometry and of the structure of a huge majority of molecules. These two complementary models assume the hypothesis of localized groups of electrons in other words stable numbers of electrons can be found in defined regions of a molecule. Similarly, the commonly admitted picture represents isolated atoms as composed of successive concentric shells encompassing the nucleus, each containing a definite number of electrons given by the aufbau rules.

The hypothesis of localized groups of electron suggests that the space occupied by the isolated atom or the molecule can be divided in parts associated with each group. These parts will be hereafter called electron group localization regions. Consider any set of M space-filling non overlapping volumes V_A. The electron population of these volumes,

$$\bar{N}_A = \int_{V_A} \rho(\mathbf{r}) d\mathbf{r}$$

is the integral over V_A of the one-electron density of probability function $\rho(\mathbf{r})$. $\bar{N}_A$ can be interpreted as of the averaged number of electrons visiting the volume V_A obtained after an infinite number of electron counts. Each individual count yields a whole number in the range $[0,N]$, N being the number of electrons of the system, therefore $\bar{N}_A$ can be rewritten as the weighted sum:

$$\bar{N}_A = \sum_0^N i p_i$$

where p_i is the probability of finding i and only i electrons within V_A. A correspondence with chemical objects such as bonds, atomic shells, lone pairs, functional groups can be established once the boundaries of the volumes are such as they minimize the fluctuations of the electron count. The present method aims to build a tool enabling the determination of "good" boundaries with the help of elementary statistical methods or information theory. The estimation of the electron population uncertainties and their further deeper analysis require the evaluation of the variances, $\sigma^(N_A)$, of related quantities such as the index of dispersion, $D(N_A) = \sigma^(N_A)/\bar{N}_A$, the coefficients of variation or relative uncertainties, $c_v = \sigma(N_A)/\bar{N}_A)$, and of the covariance matrix. The index of dispersion , also called variance to the mean ratio, Fano factor or relative fluctuation, characterizes the data distributions: $D < 1$ and $D > 1$ correspond to under- and over-dispersed regimes whereas the Poisson distribution yields $D = 0$. On the information theory side, the mutual information[17,18], MI, or rate of transmission, measures the dependence of two random variables between them. In the case of N electrons distributed over M volumes V_A, the expression of $-MI$ is:

$$MI = \sum_A \sum_B p_{AB}(1,2) \log \frac{P_{AB}(1,2)}{P_A(1) p_B(2)} \tag{1}$$

where $p_A(1)$ is the probability of finding a given electron (1) in V_A, $p_B(2)$ the probability of finding a given electron (2) in V_B and $p_{AB}(1,2)$ the probability of the joint event, a given electron (1) in V_A and another given electron (2) in V_B. The mutual information has the following properties:

1. For independent distributions $MI = 0.0$
2. For any partition $MI \geq 0$.
3. For any N-electron system, the upper bound of MI is provided by the continuous distribution limit.
4. For a partition into k localization regions, the best boundaries are those which yield the maximum value of MI, $MI_{max}(k)$.

Since electrons are quantum particles, the different statistical quantities mentioned above should be observables or rely on observables. As already mentioned $\bar{N}_A$ being a measure it can be expressed as the expectation value of an operator. Such an operator, the count operator, $\hat{N}_A$, has been introduced almost half a century ago by Diner and Claverie[19]:

$$\hat{N}_A = \sum_i^N \hat{y}(\mathbf{r}_i) \quad \text{with} \quad \hat{y}(\mathbf{r}_i) \begin{cases} \hat{y}(\mathbf{r}_i) = 1 & \mathbf{r}_i \in V_A \\ \hat{y}(\mathbf{r}_i) = 0 & \mathbf{r}_i \notin V_A \end{cases} \tag{2}$$

In Eq. 2, $\mathbf{r}_i$ is the position vector of the electron labelled by i and N is the total number of electrons in the investigated system. The eigenvalues of $\hat{N}$ are the set of whole numbers $0,\ldots N$. The uncertainty can be further evaluated by considering the variance, $\sigma^2(\bar{N}_A)$, for which an operator can be derived from Eq. 2 applying the correspondence principle. In a paper published in 2004[20], these ideas were reformulated in the framework of the multivariate analysis of the variance in which the covariance matrix elements provide a measure of the delocalization between pairs of regions. For this purpose the vectorial population operator of dimension M is introduced together with the corresponding eigenvalue and expectation value vectors:

$$\hat{\boldsymbol{N}} = \begin{pmatrix} \hat{N}_A \\ \vdots \\ \hat{N}_M \end{pmatrix} \boldsymbol{N} = \begin{pmatrix} N_A \\ \vdots \\ N_M \end{pmatrix} \bar{\boldsymbol{N}} = \begin{pmatrix} \bar{N}_A \\ \vdots \\ \bar{N}_M \end{pmatrix} \tag{3}$$

The trivial sum rules

$$\sum_M \hat{N}_A = \sum_M N_A = \sum_M \bar{N}_A = N \tag{4}$$

are the consequence of the space-filling non overlapping requirement. The covariance operator $\widehat{\boldsymbol{\sigma}^2}$ is a matrix operator whose elements are deduced from their classical expression by the correspondence principle:

$$\widehat{\boldsymbol{\sigma}^2_{AB}} = \hat{N}_A \hat{N}_B - \bar{N}_A \bar{N}_B \tag{5}$$

It is the product of the column vector $(\hat{\mathbf{N}}-\bar{\mathbf{N}})$ by its transposed, i.e.:

$$\widehat{\boldsymbol{\sigma}^{\mathbf{2}}} = (\hat{\mathbf{N}}-\bar{\mathbf{N}})(\hat{\mathbf{N}}-\bar{\mathbf{N}})^{\dagger} \tag{6}$$

The product of the population operators appearing in eq. 5 can be expressed in terms of pair population operators, i.e.:

$$\hat{N}_A\hat{N}_B = \sum_i^N\sum_j^N \hat{y}(\mathbf{r}_i)\hat{y}(\mathbf{r}_j) = \hat{\Pi}(V_A,V_B) + \delta_{AB}\hat{N}(\Omega_A) \tag{7}$$

where $\hat{\Pi}(V_A,V_B)$ is the spin free pair density operator:

$$\hat{\Pi}(V_A,V_B) = \sum_i^N\sum_{j\neq i}^N \hat{y}(\mathbf{r}_i)\hat{y}(\mathbf{r}_j) \text{ with } \begin{cases} \hat{y}(\mathbf{r}_i)=1 & \mathbf{r}_i\in V_A \\ \hat{y}(\mathbf{r}_j=1 & \mathbf{r}_j\in V_B \\ \hat{y}(\mathbf{r}_i)=0 & \mathbf{r}_i\notin V_A \\ \hat{y}(\mathbf{r}_j=0 & \mathbf{r}_j\notin V_B \end{cases} \tag{8}$$

Therefore, the covariance matrix is an observable, the elements of which can be evaluated from the one- and two-electron density of probability functions, i.e.:

$$\begin{aligned} \langle\widehat{\boldsymbol{\sigma}^{\mathbf{2}}_{AB}}\rangle &= \bar{\Pi}_{AB} - \bar{N}_A(\bar{N}_B-\delta_{AB}) \\ &= \int_{V_A}\int_{V_B} \Pi(\mathbf{r},\mathbf{r}')d\mathbf{r}d\mathbf{r}' - \bar{N}_A(\bar{N}_B-\delta_{AB}) \end{aligned} \tag{9}$$

They appear to be the difference beween the "quantum mechanical" pair populations, $\bar{\Pi}_{AB}$ and their "classical" analogs $\bar{N}_A\bar{N}_B$ or $\bar{N}_A^2-\bar{N}_A$ in the case of the diagonal elements.

The probabilities involved in the expression of *MI* are proportional to $\bar{N}_A$ and Π_{AB}, respectively:

$$p_A(1) = \frac{\bar{N}_A}{N}, \qquad p_{AB}(1,2) = \frac{\bar{\Pi}_{AB}}{N(N-1)} \tag{10}$$

Moreover, $p_{AB}(1,2)$ can be rewritten as:

$$\begin{aligned} p_{AB}(1,2) &= \frac{1}{N(N-1)}\int_{V_A}\int_{V_B}\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\left(1+h_{xc}(\mathbf{r}_1,\mathbf{r}_2)\right)d\mathbf{r}_1d\mathbf{r}_2 \\ &= \frac{N}{N-1}p_A(1)p_B(2) + \int_{V_A}\int_{V_B}\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)h_{xc}(\mathbf{r}_1,\mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2 \end{aligned} \tag{11}$$

where $h_{xc}(\mathbf{r}_1,\mathbf{r}_2)$ is the exchange correlation hole, therefore *MI* measures the inference of the exchange-correlation interaction on the distribution of the electron density over the localization regions.

The determination of the localization regions is achieved with the help of an objective function which provides a global measure of the localization (or delocalization) of the electron groups. This function, $F_k(V_A,V_B,\dots)$, depends of the number of localization volumes involved in the partition and of their boundaries. The choice of the objective function should also be guided by a simplicity requirement. As candidates in the statistical analysis approach we have considered the Froebenius norm of the covariance matrix, $\|\boldsymbol{\sigma}^{\mathbf{2}}\|_F$, the sum of the indexes of dispersion (ΣD) $\Sigma D = \sum_{V_A}\sigma^2(\bar{N}_A)/\bar{N}_A$, the sum of the coefficients of variation or relative uncertainties (Σc_v) $\Sigma c_v = \sum_{V_A}\sigma(\bar{N}_A)/\bar{N}_A$ and the mutual information in the information theory perspective.

The boundaries of the localization regions are simultaneously determined by minimizing the objective function with respect to the bounding surfaces defining the localization volumes starting from a guessed partition. In the case of the mutual information $-MI$ is the minimized function. The optimization scheme and the construction of the guess both depend on the nature of the investigated chemical objects.

It is important to note that the approach described here is totally independent of i) the way the one and two-electron densities have been determined, ii) external references such as promolecular densities or iii) functions explicitly relying on approximations. Its application only require the availability of the one and two-electron densities.

III. ATOMIC SHELLS.

The determination of non overlapping atomic shell structures has been the subject of an abundant literature reviewed in the introduction of Miroslav Kohout's article introducing the curvature of the electron position uncertainty (PUC) as localization indicator[12]. Most functions described in this review usually display an atomic shell structure but only few of them yield shell populations close to the occupancies expected from the *aufbau* principle although discrepancies may occur for transition elements. For example, PUC does not display the valence shells of Ru and Rh whereas it shows an extra external shell for Ag.

In a first step, the partitioning ability of the objective functions has been investigated on single boundary searches. As shown in Table I, $\|\boldsymbol{\sigma}^{\mathbf{2}}\|_F$ and Σc_v fail to find the shell boundaries of some atoms whereas ΣD and $-MI$ are the only objective functions yielding the expected results along the whole series of elements. The Froebenius norm and Σc_v present failures for every boundary, for example they do not display the valence shells of the Cr-Zn transition elements. The index of dispersion of the integrated density beyond the outer radius works nicely for light elements but displays a shoulder rather than a minimum for Nb and Mo[11]; adding the index of dispersion of the integrated density within this radius density gives rise to a minimum.

TABLE I. Atoms for which $\|\boldsymbol{\sigma}^{\mathbf{2}}\|_F$ and Σc_v fail to find the expected shell boundary.

boundary	$\|\boldsymbol{\sigma}^{\mathbf{2}}\|_F$	Σc_v
K-L	Rb-Rn	Mn-Rn
L-M	Tl-Rn	Na-Lu, Re, Ir
M-N	Cr-Zn	Cr-Zn
N-O	Zr-Rh, Ag-Sn Tm	Y-Sb, Er
O-P	Sm, Dy, Tm, Hf-Bi	Cs, La, Pr-Gd, Dy, Er-Rn

In a second step, the shell radii are determined by optimizing the objective function involving all the shells expected

from the Aufbau principle. The search is carried out with the Fletcher and Powell algorithm[21] starting from the radii determined in the previous step. For all elements convergence is achieved for both functions. The complete set of results, i.e. the shell radii, shell populations and variance of the population is available as supplementary information.

a. Core shells The core shell populations are closer to the number of electrons expected from occupancies than those calculated by other methods. The K-shell populations range in the intervals [2.01-2.08] for ΣD and [2.01-2.05] for $-MI$ whereas the deviations from 2.0 observed on smaller samples may attain 0.2 for *ELF*[9] and 0.3 for PUC[12]. The index of dispersion increases with the atomic number reaching 0.33 for the sixth period elements. L shell populations are calculated by both methods around 8, 8.4±0.3,for Ne-Rn element with an index of dispersion increasing from 0.07 to 0.34. The M and N shell populations main group elements are also in reasonable agreement with the expectations: $17.8 \pm 0.3, 17.85 \pm 0.25$ (M), $30.25 \pm 0.4, 30.43 \pm 0.4$ (M) for ΣD and $-MI$ respectively. However, the discrepancies may be as large as 3.0 for transition and rare earth elements.

b. valence shells Valence shells populations of main group elements are always calculated in close agreement with the electron configuration occupancies, the differences do not exceed 0.3. The populations and variance of the valence shell of periods 4-6 transition elements are presented in Table II. Unlike *ELF*[9] and PUC[12] which yield populations in good agreement with the N shell occupancy, values obtained by ΣD and $-MI$ for Mn-Ni atoms are less by c.a. 0.8 than the expectations. As a general rule the descriptions provided by ΣD and $-MI$ for early and late transition metals are consistent with their ground state configurations. Deviations larger than 0.5 occur for Nb, Mo and Ru in period 5 and for all the period 6 transition elements except Ta, Pt, Au and Hg. In the rare earth series, the averaged P shell population is 1.86. However, populations less than 1.0 are calculated for Sm and Er. The indexes of dispersion of the valence populations of transition and rare earth elements, ~ 0.5, are about twice those calculated for the main group elements of the same period.

The distribution of the probabilities of finding n electrons in a given localization region provides a complementary picture characterizing the calculated shell structure[22,23] and be further used to support chemical interpretations. The bar chart displayed in figure 1 shows maximum probabilities for $n = 8, 17$ and 1 for the L, M and N shells, respectively. Both distributions present a slight asymmetry. The accessible electron numbers of the M shell correlate with the oxidation states (OS), i.e OS=$18 - n$.

TABLE II. Transition elements: valence shell Aufbau occupancy n, populations $\bar{N}$ and population variance $\sigma^2(\bar{N})$ calculated by ΣD and $-MI$.

		ΣD		$-MI$	
		$\bar{N}$	$\sigma^2(\bar{N})$	$\bar{N}$	$\sigma^2(\bar{N})$
Sc	2	2·08	0·49	2·04	0·49
Ti	2	2·11	0·56	2·06	0·56
V	2	2·11	0·60	2·05	0·59
Cr	1	1·23	0·61	1·14	0·57
Mn	2	1·27	0·64	1·17	0·59
Fe	2	1·47	0·77	1·39	0·73
Co	2	1·16	0·59	1·05	0·54
Ni	2	1·13	0·58	1·02	0·52
Cu	1	1·09	0·55	0·97	0·49
Zn	2	2·08	0·76	2·01	0·74
Y	2	2·23	0·64	2·19	0·63
Zr	2	2·40	0·79	2·35	0·78
Nb	1	2·05	1·06	1·99	1·03
Mo	1	3·57	1·72	3·62	1·74
Tc	2	1·69	0·98	1·93	0·88
Ru	1	1·57	0·93	1·48	0·88
Rh	1	1·41	0·84	1·30	0·78
Pd	0				
Ag	1	1·14	0·68	0·94	0·57
Cd	2	2·18	1·0	2·10	0·97
Hf	2	2·67	0·93	2·67	0·93
Ta	2	2·32	0·88	2·30	0·88
W	2	2·89	1·18	2·89	1·18
Re	2	0·79	0·44	0·65	0·34
Os	2	1·03	0·63	0·76	0·45
Ir	2	1·02	0·62	0·79	0·46
Pt	1	0·91	0·55	0·68	0·40
Au	1	0·74	0·36	0·62	0·31
Hg	2	1·94	0·92	1·99	0·94

IV. MOLECULES

The optimization of the localization regions in small molecules has been achieved by an iterative trial and error algorithm. The molecular volume is represented by a rectangular prism enclosing more than 99.5% of the charge density and divided in cubic cells of 1.0^{-3} bohr^3. Starting from a guessed partition provided by *ELF* , each cell is assigned to a localization region. For each cell at the border of two regions, the assignment is changed and the objective function recalculated. If the value of the objective function is improved, the new assignment is kept and otherwise rejected. The process is repeated until convergence is achieved.

Table III compares the localization region populations calculated for a series of small molecules. The nomenclature adopted to label the localization regions is that introduced for *ELF* attractors and basins a quarter of century ago[24]. Almost all the values calculated with ΣD and $-MI$ are close one another and in good agreement with those predicted by the Lewis structure. The discrepancies with respect to the expected whole numbers being less than 0.25, except for CO and N_2. The *ELF* partition always yields larger populations for lone pairs, V(A), at the expense of bonds, V(A, B), as for example in ammonia, water, hydrogen fluoride and hydrogen chloride molecules where the V(A,H) populations nicely correlates with the $\chi(\mathrm{A}) - \chi(\mathrm{H})$ electronegativity difference. In linear molecules, when the axial symmetry merges several formal lone pairs in a single region, the population of the bond becomes unexpectedly large with ΣD and $-MI$. For exam-

TABLE III. Populations of localization regions of AB_x and C_2H_{2n} ($n = 1,2,3$) molecules determined with ΣD, $-MI$ and ELF.

Molecule		C(A)			C(B)			V(A)			V(B)			V(A,B)	
	ΣD	-MI	ELF	ΣD	-MI	ELF	ΣD	-MI	ELF	ΣD	-MI	ELF	ΣD	-MI	ELF
LiH	2·01	2·01	2·01										1·99	1·99	1·99
BH_3	2·05	2·06	2·06										1·99	1·99	1·99
CH_4	2·06	2·05	2·09										1·99	1·99	1·98
NH_3	2·07	2·05	2·11				2·03	2·05	2·16				1·97	1·97	1·91
H_2O	2·07	2·04	2·12				2·07	2·09	2·27				1·89	1·89	1·67
HF	2·08	2·05	2·13				5·70	5·76	6·42				2·21	2·18	1·45
HCl	10·06	10·05	10·07				5·96	6·03	6·18				1·98	1·92	1·75
FHF^-	2·08	2·06	2·10				7·70	7·77	7·77	0·40	0·32	0·26			
LiF	2·03	2·03	2·03	2·08	2·08	2·11				7·88	7·88	7·85			
CO	2·04	2·02	2·08	2·10	2·05	2·12	2·19	2·19	2·24	2·88	2·93	4·02	4·86	4·69	3·18
N_2	2·05	2·03	2·10				2·42	2·41	3·21				5·05	5·12	3·39
F_2	2·09	2·06	2·21				5·88	6·13	6·66				2·01	1·59	$2\times0·13$
		C(C)			V(C,H)				V(C,C)						
C_2H_2	2·05	2·02	2·10	2·10	2·02	2·32		5·70		5·91		5·14			
C_2H_4	2·05	2·03	2·09	1·99	1·99	2·11		$2\times1·99$		$2\times2·0$		$2\times1·69$			
C_2H_6	2·06	2·04	2·09	1·96	1·95	2·0		2·14		2·21		1·81			

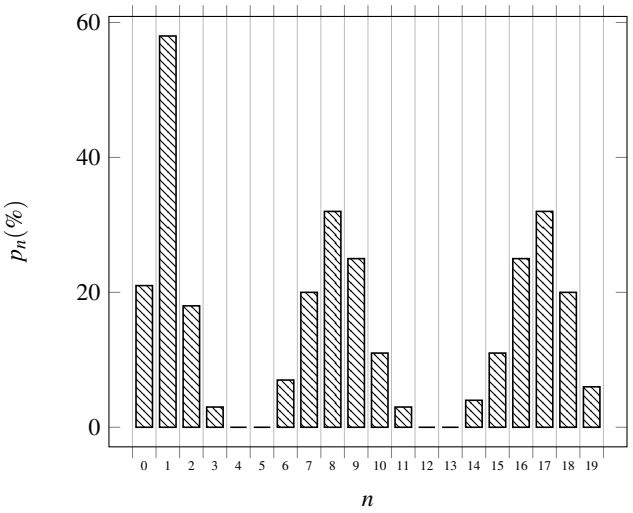

FIG. 1. Probability distribution of electron numbers in the N, M, L shells of Ni

ple, the V(F,H) population of hydrogen fluoride amounts to 2.21 and 2.18 for ΣD and $-MI$, respectively. We suspect the large $\bar{N}_{V(F)}$ and $p_{V(F)}$ appearing in denominators in the actual expression of these objective functions to be responsible for this uncomfortable feature. The Froebenius norm and Σc_v, do not work better yielding 0.0 and 3.49, respectively, values which support our explanation. Attempts to divide V(F) in three subregions are unsuccessful. However the bonding picture of FHF^- provided by ΣD and $-MI$ does not noticeably differs from the ELF one as the V(H) population is less than 0.5 for the three approaches.

Another interesting example of discrepancies between ΣD and $-MI$ population happens for the difluorine molecule. The description of the valence density given by ELF shows a bond region split in two small basins, each with a very low population. This led Rosa LLusar and co-workers to call this pattern "protocovalent"[25]. Moreover the off diagonal covariance matrix element $\sigma^2(V_F,V_F)$ has a rather large absolute value indicating a significant delocalization between the lone pairs of the two fluorine atoms. Weak bond populations and large absolute value of covariance between lone pairs of adjacent atoms is the ELF signature of charge-shift bonding[26,27]. The picture provided by the covariance matrix Froenenius norm minimization fully agree with ELF. Since the fluorine lone pairs are merged by the axial symmetry, ΣD and $-MI$ favour perfect pairing in the bond region: during the optimization process not only the two "protocovalent" moieties of the F−F bond are merged and the population of the resulting region increased to 2.01(ΣD) and 1.59($-MI$) but also $\sigma^2(V_F,V_F)$ is halved.

In the bonding picture of CO and N_2 provided by ΣD and $-MI$, the populations of the bond exceeds its ELF value by more than 1.0. Figure 2 as well as the population values show that the increase of the bond at the expense of the oxygen lone pairs is the main difference between $-MI$ and ELF. Whereas, ELF provides a description of CO in terms of superposition of Lewis diagrams dominated by the $|C{=}O\rangle$ and $\langle C{-}O\rangle$ structures[28], the MI results is explained by considering $|C{=}O\rangle$ and $|C{\equiv}O|$ as major contributions. The bond population of N_2 calculated with ΣD and $-MI$ is intermediate between a triple bond and a double bond while that obtained from the ELF partition has obviously a rather large single bond contribution.

In the case of the hydrocarbons all method yield populations close to the Lewis diagram expectations. In agreement with the ELF findings, the C_2H_4 double bond gives rise to two distinct pair regions whereas in C_2H_2 the $D_{\infty h}$ merges the three bonding pairs in a single region.

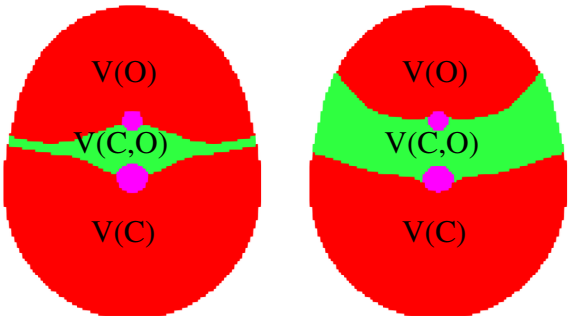

FIG. 2. From left to right ELF and $-MI$ localization regions of CO in a σ_v plane. Colour code: cores: magenta, lone pairs: red, bond: green.

V. CONCLUSION

The analysis of the charge density presented here provides a picture in terms of localized electron groups close to the chemical representation derived from Lewis's valence theory. The approach used in this article is inspired by the data processing of experimental results since the statistical or information theory analysis is carried out on the output of electron count numerical experiments. As one and two-electron density are observables, the same treatment can be applied in principle to experimental data. We endeavoured to restrict the arbitrary factors to the choice of the objective function.

In light of these preliminary results, the ΣD and $-MI$ are the objective functions which yield, at least qualitatively, the best agreement with the chemical models for both atomic shell structure and "electron shared" interactions[2,4] in molecules. In the case of ionic bonding (LiF and LiH) the cation and ionic moieties are perfectly separated. From a quantitative point of view, the populations of the valence regions calculated with these methods are usually closer to those deduced from classical structural formulas than those provided by the ELF analysis. It is particularly the case of the ELF valence populations of H_2O and N_2 for which the difference with the perfect pairing model were interpreted as unphysical rather than as a consequence of bond polarity or mesomery.[29,30]. Attempts to decompose groups formally containing several electron pairs into pair subgroups failed for atoms and for linear molecules exemplifying the importance of the nucleus-electron potential (the external potential in the DFT vocabulary) whose symmetry constraints the number of distinct electron groups as it has been also shown with the ELF. It is interesting to note that multivariate statistics ($\sum D$) and information theory ($-MI$) yield very similar results.

The Maximum Probability Domains (MPDs)[11,31–33] is an other, and even better approach attesting the reality of spatially localized electron groups. Unfortunately the regions generated by this method are not constrained to be non-overlapping, and therefore the development of a population analysis from MPDs appears difficult. On the global objective function optimization side, all the features of the ELF population analysis can be adapted in a straightforward fashion and most concepts introduced in this framework remain valid. However, it is no longer rigorous to use the vocabulary and the mathematical properties of dynamical systems.

Although ΣD and $-MI$ often yield very close results, it should be interesting to consider other objective functions, such $\|\boldsymbol{\sigma^2}\|_F$ and Σc_v, in parallel studies aiming to understand the evolution of a chemical property along a series of parent molecules. In such studies, it is further possible to bring together regions corresponding to a chemical subset of the system in order to simplify the chemical explanations, as for example the proton donnor and proton acceptor moieties of an hydrogen bond complex[34].

Beyond the quantitative analysis of the charge density, the method described in this article should be used to conceptually refine our chemical representation of the matter. The variance and information related criteria are rooted on the statistical interpretation of quantum mechanics. They enables to corroborate or falsify the hypothesis of localized electron groups by numerical experiments providing precise, and to some extent mathematizable, definitions of the related chemical objects. To this respect, our method intend to contribute to a clarification of the weakly defined concepts of the chemical bonding theory, and among them the concept of bond itself, in order to improve the scientificity of interpretative computational chemistry.

Computational details

a. Atoms

- Method: restricted Hartree Fock
- Periods 2-5 elements: quadruple zeta basis set[35,36]
- Period 6 elements: triple zeta basis set[37–39]

b. Molecules

- B3LYP[40–43]
- Basis set: 6-311G(d,p)[44,45]
- Nuclear coordinate of the optimized geometry.

c. Softwares

- *ab initio calculations* Gaussian09[46],
- ELF partition TopMod[47,48]

Supplementary material

A tribute celebrating Klaus Ruedenberg, formula enabling the calculation of partial overlap integrals of cartesian gaussian type orbitals and the complete $\sum D$ and MI Li-Rn shell structure tables are reported in the Supplementary Material.

AVAILABILITY OF DATA	STATEMENT OF DATA AVAILABILITY
Data available on request from the author	The data that support the findings of this study are available from the corresponding author upon reasonable request. The dedicated programs written by the authors are also available on request.

[1] S. Alvarez, R. Hoffmann, and C. Mealli, "A bonding quandary–or–a demonstration of the fact that scientists are not born with logic," Chem. Eur. J. **15**, 8358–8373 (2009).

[2] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory* (Oxford Univ. Press, Oxford, 1990).

[3] A. D. Becke and K. E. Edgecombe, "A simple mesure of electron localization in atomic and molecular systems," J. Chem. Phys. **92**, 5397–5403 (1990).

[4] B. Silvi and A. Savin, "Classification of chemical bonds based on topological analysis of electron localization function," Nature **371**, 683–686 (1994).

[5] Z. Shi and R. J. Boyd, "The shell structure of atoms and the laplacian of the charge density," J. Chem. Phys. **88**, 4375–4377 (1988).

[6] R. P. Sagar, A. C. T. Ku, J. Vedene H. Smith, and A. M. Simas, "The laplacian of the charge density and its relationship to the shell structure of atoms and ions," J. Chem. Phys. **88**, 4367–4374 (1988).

[7] M. Kohout, A. Savin, and H. Preuss, "Contribution to the electron distribution analysis. i. shell structure of atoms," J. Chem. Phys. **95**, 1928–1942 (1991).

[8] K. Sen, M. Slamet, and V. Sahni, "Atomic shell structure in hartree—fock theory," Chem. Phys. Lett. **205**, 313 – 316 (1993).

[9] M. Kohout and A. Savin, "atomic shell structure and electron numbers," Int. J. Quant. Chem. **60**, 875–882 (1996).

[10] M. Kohout, "Occupation numbers for atomic shells in direct space bounded by the maxima of the one-electron potential," Int. J. Quant. Chem. **83**, 324–331 (2001).

[11] A. Savin, "Probability distributions and valence shells in atoms," in *Reviews of Modern Quantum Chemistry: A Celebration of the Contributions of Robert G.Parr*, Vol. I, edited by K. D. Sen (World Scientific, Singapore, 2002) pp. 43–62.

[12] M. Kohout, "Atomic shell structure determined by the curvature of the electron position uncertainty," Mol. Phys. **114**, 1297–1303 (2016).

[13] G. N. Lewis, "The atom and the molecule," J. Am. Chem. Soc. **38**, 762–786 (1916).

[14] G. N. Lewis, *Valence and the Structure of Atoms and Molecules*, Am. Chem. Soc. Monograph Series (The Chemical Catalog Company, New York, 1923) p. 163.

[15] R. J. Gillespie and R. S. Nyholm, "Inorganic stereochemistry," Quart. Rev. Chem. Soc. **11**, 339–380 (1957).

[16] R. J. Gillespie, *Molecular Geometry* (Van Nostrand Reinhold, London, 1972).

[17] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (The University of Illinois Press, Urbana, 1949).

[18] T. M. Cover and J. A. Thoms, *Elements of information theory*, 2nd ed. (Wiley Interscience, New York, 2006).

[19] S. Diner and P. Claverie, "Statistical and stochastic aspects of the delocalization problem in quantum mechanics," in *Localization and Delocalization in Quantum Chemistry*, Vol. II, edited by O. Chalvet, R. Daudel, S. Diner, and J. P. Malrieu (Reidel, Dordrecht, 1976) pp. 395–448.

[20] B. Silvi, "How topological partitions of the electron distributions reveal delocalization," Phys. Chem. Chem. Phys. **6**, 256–260 (2004).

[21] R. Fletcher and M. Powell, "A rapid convergent descent method for minimization," The Comput. J. **6**, 163–168 (1963).

[22] E. Chamorro, P. Fuentalba, and A. Savin, "Electron probability distribution in AIM and ELF basins," J. Comput. Chem. **24**, 496–504 (2003).

[23] E. Cancès, R. Keriven, F. Lodier, and A. Savin, "How electrons guard the space: shape optimization with probability distribution criteria," Theor. Chem. Acc. **111**, 373–380 (2004).

[24] A. Savin, B. Silvi, and F. Colonna, "Topological analysis of the electron localization function applied to delocalized bonds," Can. J. Chem. **74**, 1088–1096 (1996).

[25] R. Llusar, A. Beltrán, J. Andrés, S. Noury, and B. Silvi, "A topological analysis of electron density in depleted homopolar chemical bonds." J. Comput. Chem. **20**, 1517–1526 (1999).

[26] S. S. Shaik, D. Danovich, B. Silvi, D. Lauvergnat, and P. Hiberty, "Charge-shift bonding–a class of electron pair bonds emerges from valence bond theory and supported by electron localization function approach," Chem. Eur. J. **21**, 6358–6371 (2005).

[27] S. S. Shaik, D. Danovich, W. Wu, and P. Hiberty, "Charge-shift bonding and its manifestations in chemistry," Nature Chemistry **1**, 443–449 (2009).

[28] C. Lepetit, B. Silvi, and R. Chauvin, "ELF analysis of out-of-plane aromaticity and in-plane homoaromaticity in carbo[n]annulenes and [n]pericyclines," J. Phys. Chem. A **107**, 464–473 (2003).

[29] R. Ponec and J. Chaves, "Electron pairing and chemical bonds: Pair localization in ELF domains from the analysis of domain averaged Fermi holes," J. Comput. Chem. **28**, 109–116 (2007).

[30] J. F. Gonthier, S. N. Steinmann, M. D. Wodrich, and C. Corminboeuf, "Quantification of "fuzzy" chemical concepts: a computational perspective," Chem. Soc. Rev. **41**, 4671–4687 (2012).

[31] M. Causà and A. Savin, "Maximum probability domains in crystals: The rock-salt structure," J. Phys. Chem. A **115**, 13139–13148 (2011).

[32] O. M. Lopes, B. Braïda, M. Causà, and A. Savin, "Advances in the theory of quantum systems in chemistry and physics," (Springer Netherlands, Dordrecht, 2012) Chap. Understanding Maximum Probability Domains with Simple Models, pp. 173–184.

[33] M. Causà, M. D'Amore, C. Garzillo, F. Gentile, and A. Savin, "The bond analysis techniques (elf and maximum probability domains) application to a family of models relevant to bio-inorganic chemistry," in *Applications of Density Functional Theory to Biological and Bioinorganic Chemistry*, edited by M. V. Putz and D. M. P. Mingos (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013) pp. 119–141.

[34] B. Silvi and H. Ratajczak, "Hydrogen bonding and delocalization in the *ELF* analysis approach." Phys. Chem. Chem.Phys. **18**, 27442–27449 (2016).

[35] F. Weigend, F. Furche, and R. Ahlrichs, "Gaussian basis sets of quadruple zeta valence quality for atoms h-kr," J. Chem. Phys. **119**, 12753–12762 (2003).

[36] G. A. Ceolin, R. C. de Berrêdo, and F. E. Jorge, "Gaussian basis sets of quadruple zeta quality for potassium through xenon: application in ccsd(t) atomic and molecular property calculations," Theor. Chem. Acc. **132**, 1339 (2013).

[37] C. T. Campos and F. E. Jorge, "Triple zeta quality basis sets for atoms rb through xe: application in ccsd(t) atomic and molecular property calculations," Mol. Phys. **111**, 167–173 (2013).

[38] L. S. C. Martins, F. E. Jorge, and S. F. Machado, "All-electron segmented contraction basis sets of triple zeta valence quality for the fifth-row elements," Mol. Phys. **113**, 3578–3586 (2015).

[39] A. Z. de Oliveira, I. B. Ferreira, C. T. Campos, F. E. Jorge, and P. A. Fantin, "Segmented all-electron basis sets of triple zeta quality for the lanthanides: application to structure calculations of lanthanide monoxides," J. Mol. Model. **25**, 38 (2019).

[40] A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," J. Chem. Phys. **98**, 5648–5652 (1993).

[41] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Phys. Rev. A **38**, 3098–3100 (1988).

[42] C. Lee, Y. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," Phys. Rev. **B37**, 785 (1988).

[43] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, "Results obtained with the correlation energy density functionals of becke and lee, yang and parr," Chem. Phys. Lett. **157**, 200 (1989).

[44] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, "Self-consistent molecular orbital methods. xx. a basis set for correlated wave functions," J. Chem. Phys. **72**, 650–654 (1980).

[45] M. J. Frisch, J. A. Pople, and J. S. Binkley, "Self-consistent molecular-orbital methods. xxv: Supplementary functions for gaussian basis sets." J. Chem. Phys. **80**, 3265–3269 (1984).

[46] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson,

H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö.. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 Revision D.01," (2009), gaussian Inc. Wallingford CT 2009.

[47] S. Noury, X. Krokidis, F. Fuster, and B. Silvi, "Computational tools for the electron localization function topological analysis." Comput. in Chem. **23**, 597–604 (1999).

[48] S. Noury, X. Krokidis, F. Fuster, and B. Silvi, "Topmod package," (1997).